Threshold interaural time differences and the centroid model of sound localization

William M. Hartmann* and Andrew Brughera

*Corresponding author’s address: Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, hartmann@pa.msu.edu

The centroid display model of sound lateralization hypothesizes a two-dimensional array of brain-stem cells with wide ranges of best frequencies ($f_c$) and best interaural time delays (ITD, $\tau$). The cells are distributed according to a cell population density function $p(f_c, \tau)$, and images are lateralized according to the centroid of an excitation pattern on this array, the rate-ITD function, weighted by $p(f_c, \tau)$. The ability of the centroid display to predict the frequency dependence of human ITD thresholds was tested by calculations using model cells for the medial superior olive, as the origin of the rate-ITD function. The cells had synaptic inputs, membrane channel dynamics, and time constants established by physiological measurements. Cells were driven with realistic frequency-dependent synchrony. The centroid model could successfully account for the high-frequency human data, but greatly underestimated thresholds for lower frequencies. Similar calculations with a rate-difference model was successful in the lower frequency range but failed at high frequencies. Therefore, we propose a hybrid display model consisting of a centroid code at high frequencies and a rate code at low. [Supported by the AFOSR and NIDCD.]
INTRODUCTION

Neural encoding of the interaural time difference (ITD) is a major component of sound localization (Strutt, 1907), which also enhances speech reception in noisy environments. The present article is concerned with the threshold for detecting ITD in sine tones, particularly its unusual dependence on the frequency of the tone. It has been known for years that ITD thresholds grow extremely rapidly as the tone frequency approaches 1500 Hz (e.g. Zwislocki and Feldman, 1956).

The goal of the present article is to model the frequency dependence of the thresholds as measured in recent experiments by Dunai et al. (2012). The measurements were done for a baseline ITD of zero (midline) for four listeners using a two-alternative forced-choice task based on the listener’s sensation of laterality — left or right. The stimulus ITD was varied using an adaptive three-down one-up staircase method. The experiments made a careful study of the high-frequency behavior of the thresholds and showed that they grew faster than exponentially with increasing frequency until it became impossible to measure thresholds at 1450 Hz or above. The experimental data appear as the open symbols in Fig.2.

The modeling consists of two stages. The first stage is a physiologically realistic computational model of individual brainstem neurons where ITD sensitivity is derived using a binaural coincidence detector mechanism. In mammals, this mechanism occurs in principal cells of the medial superior olive (MSO), which incorporate neural inputs that are phase-locked to sound at the left and right ears, and are tuned in both frequency (f) and in best delay (τ) (Yin and Chan, 1990; Goldberg and Brown, 1969).

The second stage of modeling, which is explored in the greatest detail here, is the binaural display — a process by which information from an array of MSO cells is extracted to enable spatial localization of the tone. Two displays were considered: a centroid-based place model and a rate-difference model. The centroid model is an extension of the Jeffress (1948) cross-correlator — postulating a wide two-dimensional array tuned in frequency and best delay. The rate-difference model postulates a simpler, reduced array. The present treatment begins with the MSO cell itself, used as reliable input to test both of the binaural display models.

MSO CELL MODEL

A physiologically-based ITD-sensitive MSO neuron model (Dunai et al., 2012) was developed in which large action potentials are limited to the axon with only minor back-propagation to the soma, similar to real MSO neurons (Scott et al., 2007). Synaptic time constants were slower than modeled previously (Zhou et al., 2005; Brand et al., 2002). Excitatory synaptic time constants (τErise = τEdecay = 0.4) reflect recent physiological measurements in MSO neurons (Fischl et al., 2012). Inhibitory synaptic time constants are either set as “slowly decaying” with τIrise = 0.4 ms and τIdeca y = 2 ms to match measured values in MSO neurons (Fischl et al., 2012; Magnusson et al., 2005), or set as “rapidly-decaying” with τIrise = τIdeca y = 0.4 ms to produce contralateral-leading best-ITDs similar to those measured in Brand et al. (2002). The binaural displays received inputs in the form of discharge rate vs. ITD functions derived from the model MSO neurons.

Figure 1(A) shows the firing rate for an MSO cell with only excitatory inputs. The rate-
Figure 1: Discharge rate of a model MSO cell as a function of ITD in acoustic tones for six different frequencies from 250 to 1500 Hz. (A) With only excitatory inputs. (B) With excitatory inputs and slowly decaying inhibition. (C) With excitatory inputs and rapidly decaying inhibition.
ITD functions were highly modulated at and below 1000 Hz, and became progressively less modulated with increasing frequency above 1000 Hz, becoming relatively flat at 1500 Hz.

Parts B and C of Fig.1 include, respectively, slowly-decaying inhibition and rapidly-decaying inhibition. Compared with the purely excitatory condition, both types of contralateral inhibition reduced overall discharge rates.

Slowly-decaying inhibition increased the relative depth of modulation except at 500 Hz, where the modulation remained complete. At 1000 Hz and below, slowly-decaying inhibition decreased the half-width of each rate-ITD function by more than 12 percent, indicating that the sharpness of ITD tuning was increased by the inhibition: at 1250 Hz, the half-width decreased slightly; and at 1500 Hz, the half-width increased slightly.

Rapidly-decaying inhibition, across all input frequencies, shifted the best ITD from zero (in the purely excitatory condition) to contralateral-leading ITD, such that the steepest slope in each rate-ITD function occurred near the midline (zero ITD). Rapidly decaying inhibition increased the depth of modulation except at 500 and 1000 Hz, where it remained maximal. Rapidly decaying inhibition also decreased half-widths by more than 12% at 500 Hz and below, but had less effect on ITD tuning at 750 Hz and above.

**RATE-DIFFERENCE DISPLAY**

The rate-difference model was proposed for low-frequency sound localization in mammals by McAlpine et al. (2001). Unlike the Jeffress model, with an extensive array of cells, the rate-difference model can profitably be imagined to have only a single cell in the left MSO and a single cell in the right MSO for each tone frequency. A subsequent stage of processing compares the firing rates of the left and right MSO cells and obtains a measure of laterality based on that difference. For a rate-difference model, it is advantageous for the firing rate of the MSO cells to depend sensitively on the ITD. Because listeners show greatest sensitivity for small ITDs, an optimum model displaces the peak of the rate-ITD function along the best-delay axis so that a steep slope occurs in the region of small delays. Such a displacement can be obtained by adding inhibitory inputs with rapid decay to the MSO cells (Brand et al., 2002) as in part C of Fig.1. Alternatively, such a displacement may reflect delay lines of different length, or interaural differences in cochlear traveling-wave delays as modeled in Shamma et al. (1989) and Day and Semple (2011).

The predictions of the rate-difference model were compared with the human ITD thresholds by using the functions of part C of Fig. 1 to compute firing rates in computer-simulated staircase experiments that exactly followed the protocol of the human experiments. Firing rates were computed for left and right ears, and the variability in those firing rates was incorporated by adding a normally distributed deviation based on five independent computations of the MSO cell rates. Given a stimulus with an ITD that cued locations first to the left and then to the right, the simulated response was judged to be correct if the difference in excitation, right ear minus left ear, was larger on the second simulated interval than on the first. Otherwise the response was wrong. The results of the simulation are given in part (a) of Fig.2.

Figure 2 shows that the simulation was remarkably successful considering that it is free
of adjustable parameters, except for the inhibition in the MSO cell. The simulation does not agree with most of the human data at 250 Hz, but this deficiency is easily remedied by only slight changes in the excitation synaptic strength. Alternatively, assuming that the standard deviation in rate-ITD function is equal to the square root of the firing rate, divided by the square root of the tone duration, as appropriate for a Poisson process, also brings the low-frequency simulation into good agreement with experiment. The big problem with the simulation is that it does not reproduce the divergence at high frequencies that is always found for human listeners.

CENTROID DISPLAY

The centroid display model is an intuitively plausible extension of the Jeffress cross-correlation array. The hypothesis of the centroid model is that the binaural system computes a center of gravity of the excitation $E$ along the best-delay dimension $\langle \tau \rangle$ of the array, and uses that computation as an estimate of the azimuthal location of the source of a sound. Given an ITD of $\Delta t$, the centroid is

$$
\bar{\tau}(\Delta t) = \frac{\int d\tau \tau E(\tau - \Delta t)}{\int d\tau E(\tau - \Delta t)}.
$$

where the integrals are over the range of minus to plus infinity. Excitation $E(\tau) = r(\tau)p(\tau)$ is the population-weighted excitation pattern as a function of best-delay $\langle \tau \rangle$, where $r(\tau)$ is the discharge rate in MSO neurons across best-delay at the given ITD (assumed equal to the rate-ITD function of a single MSO neuron with best-delay equal to the ITD), and $p(\tau)$
is the neural population weighting-function. Because of the importance of interaural phase, function $\tau$ must depend on frequency, $f$. Function $p$ may also depend on frequency or best frequency.

The centroid display hypothesis formed the core of the position variable model invented by Stern and Colburn (1978). This model attempted to account for the lateralization of 500-Hz pure tones as measured in experiments with both interaural time differences (ITD) and interaural level differences (ILD). The position variable model was successful in modeling the results of laterality experiments, and it also accounted for data from image centering experiments that traded ITD in one direction against ILD in the other.

There are several essential ingredients that characterize centroid display models. The first is that the range of internal delays in the binaural system is long enough to register secondary maxima. For example, given a 1000-Hz tone with an ITD of 0, there is, of course, a peak along the array at a best delay of 0. There are also secondary maxima at plus and minus 1 ms, corresponding to the period of the tone. In computing the centroid, these secondary maxima are included in the integral with important consequences for the frequency dependence of the centroid, as noted by Stern and Shear (1996).

A second feature of centroid models is the weighting function that represents the density of coincidence cells (population density) $p(\tau)$. Because the array of cells is imagined to be symmetrical about the center $\tau = 0$, and because cells with brief delays are thought to be the most numerous, the population density function leads to a centrality weighting whereby cells with ITD close to zero are weighted much more strongly than cells with longer ITD. The original position variable model by Stern and Colburn used a central weighting function that was constant for small ITD and decayed exponentially for longer ITD with both short and long decay rates. That weighting function had been used by Colburn (1977) to model the masking level difference, particularly as the interaural amplitudes and phases of signal and masking noise were varied.

As noted by Stern and Shear (1996), Colburn’s $p(\tau)$ decays too slowly to successfully model the lateralization of tones at high frequency. Therefore, we investigated other similar functions for $p(\tau)$, flat for $\tau < 0.2$ ms and with a variety of decaying time constants, 0.220 ms, 0.400 ms, and 0.750 ms.

Population density $p(\tau)$ is shown in Fig. 3 by red, green and blue plots for the three values of the decay parameter $\tau_o$. The function is shown for only positive values of $\tau$, because $p(\tau)$ has even symmetry. The population density functions in Fig. 3 appear to span the range of functions used by Stern and Shear to model a frequency dependent function $p(\tau|f_o)$. The red, green, and blue functions in Fig. 3 have decay times that can be called short, medium, and long, respectively. They can be compared with Colburn’s original function shown in black. Our long-decay function, with a time constant of 0.750 ms, was chosen to correspond with Colburn’s function, which also has a long decay. From the view point of Fig. 3, we might better have chosen a time constant of 0.600 ms, the same as Colburn. Such a choice would have made our function nearly coincide with his. However, Colburn’s function has an additional very long tail, which has important effects on threshold calculation. Increasing our time constant from 0.600 to 0.750 ms models those effects better.
Figure 3: Neural population weighting functions for centroid calculations, normalized to unit area.

Adaptive Staircase Simulation

To predict centroid-display thresholds for comparison with experiment, we ran simulated adaptive staircases with procedures identical to those used in testing the rate-difference display. Centroids were computed using the excitation-only rate ITD functions shown in Fig. 1. Again, the randomness indicated by the MSO model calculations was incorporated into each value of $r(\tau)$ in the integral of Eq. 1. If the centroid for the right-leading tone was further to the right than the centroid for the left-leading tone, the response to the simulated trial was taken to be correct. Otherwise it was wrong.

Twenty simulated runs for each frequency and each $p(\tau)$ function led to the predicted regions in part (b) of Fig. 2. The colors of the model threshold plots correspond to the colors of the population density functions in Fig. 3. Figure 2 shows that the model thresholds dramatically fail to agree with the measured thresholds. The predictions at low frequency are far too low.

It is not hard to find an explanation for the failure of the centroid model. The calculated thresholds essentially depend on the ratio of the modulation in the rate-ITD function to the standard deviation as a measure of the random variations in that function. However, the centroid model involves an integral over the internal ITD axis, as weighted by $p(\tau)$. The random variations in the rate-ITD function are positive and negative with equal probability and tend to cancel in the integration. As a result, the variations have far too little effect on the calculated lateralization to agree with the experiment.

But although the centroid model fails at low frequency, it can easily reproduce the divergence at high frequency if $p(\tau)$ decays slowly enough. Our medium decay rate (0.400 ms) leads to agreement with human data at 1250 Hz and diverges at 1500 Hz. The divergence occurs because of the cancellation effect of the secondary maxima in the rate-ITD function,
which increasingly fall into the region allowed by $p(\tau)$ as the frequency increases.

CONCLUSION

Two model binaural displays were evaluated in their ability to predict human ITD discrimination thresholds for individually-presented sine tones as a function of frequency. The rate-difference display made predictions in good agreement with human threshold data for low frequencies, but failed to reproduce the divergence above 1400 Hz. By contrast, the centroid display which incorporates a place-code reproduced human thresholds above 1000 Hz, including the divergence above 1400 Hz, but failed at lower frequencies by predicting performance much better than human. Based on these results we propose that the human binaural system may operate using a hybrid display composed of a rate-difference code up to approximately 1100 Hz, and a place code at higher frequencies.

ACKNOWLEDGMENTS

We are grateful to Dr. Les Bernstein for a useful discussion about the centroid display. Mr. Zane Crawford provided valuable statistical help. This research was supported by the NIDCD grant DC-00181 and the AFOSR grant 11NL002. AB was supported by NIDCD grants DC-00100 (H.S. Colburn) and P30-DC04663 (Core Center).

REFERENCES


